Journal of Neuroscience Methods 398 (2023) 109953

Contents lists available at ScienceDirect

Journal of Neuroscience Methods

journal homepage: www.elsevier.com/locate/jneumeth

Self-attention-based convolutional neural network and time-frequency common spatial pattern for enhanced motor imagery classification

Rui Zhang, Guoyang Liu, Yiming Wen, Weidong Zhou *

School of Microelectronics, Shandong University, Jinan 250100, China

ARTICLE INFO

Keywords:
Motor imagery
Brain-computer interface
Electroencephalogram (EEG)
Self-attention-based convolutional neural network
Time-frequency common spatial pattern (TFCSP)

ABSTRACT

Background: Motor imagery (MI) based brain-computer interfaces (BCIs) have promising potentials in the field of neuro-rehabilitation. However, due to individual variations in active brain regions during MI tasks, the challenge of decoding MI EEG signals necessitates improved classification performance for practical application.
New method: This study proposes a self-attention-based Convolutional Neural Network (CNN) in conjunction with a time-frequency common spatial pattern (TFCSP) for enhanced MI classification. Due to the limited availability of training data, a data augmentation strategy is employed to expand the scale of MI EEG datasets. The self-attention-based CNN is trained to automatically extract the temporal and spatial information from EEG signals, allowing the self-attention module to select active channels by calculating EEG channel weights. TFCSP is further implemented to extract multiscale time-frequency-space features from EEG data. Finally, the EEG features derived from TFCSP are concatenated with those from the self-attention-based CNN for MI classification.
Results: The proposed method is evaluated on two publicly accessible datasets, BCI Competition IV IIa and BCI Competition III IIIa, yielding mean accuracies of 79.28 % and 86.39 %, respectively.
Conclusions: Compared with state-of-the-art methods, our approach achieves superior classification results in accuracy. Self-attention-based CNN combining with TFCSP can make full use of the time-frequency-space information of EEG, and enhance the classification performance.

1. Introduction

Brain-computer interfaces (BCIs) facilitate direct communication between the human brain and the external environment, bypassing the need for peripheral nerves and muscles (Abiri et al., 2019; Liu et al., 2022a; Xu et al., 2016). Electroencephalogram (EEG) has been widely utilized in motor imagery (MI)-based BCI, due to its non-invasive, low-cost, and portable nature (Xie et al., 2022; Zhang et al., 2021). EEGs for motor imagery tasks are obtained as subjects imagine their limbs moving without physical execution (Yu et al., 2022; Ieracitano et al., 2021). In addition, event-related desynchronization (ERD) (Graimann et al., 2010) and event-related synchronization (ERS) (Pfurtscheller et al., 2006, 1997) phenomena generated through MI-BCIs have been used to classify EEG data from a variety of tasks, thereby enabling individuals with motor disabilities to interact with their surroundings and improve their life quality (Padfield et al., 2019; Khan et al., 2020). Applications of MI-BCIs span various fields, including orthosis treatment for tetraplegic patients (Pfurtscheller and Neuper, 2001), quadcopter control (LaFleur et al., 2013), and functional mobility of exoskeletons and prostheses (Elstob and Lindo, 2016; Al-Quraishi et al., 2018).

Classification methods of EEG for MI tasks can be broadly categorized into traditional machine learning methods and deep learning methods. Traditional approaches typically involve three essential steps: preprocessing, feature extraction, and classification. Feature extraction is crucial for determining classification performance, with multiple methods employed to extract features and classify them using machine learning classifiers. For example, the common spatial pattern (CSP) is the most commonly used spatial feature extraction method for MI task classification (Koles et al., 1990; Liu et al., 2023a), discriminating EEGs from different classes by maximizing the variance for one class and minimizing it for the other. However, the conventional CSP algorithm's performance is highly dependent on the frequency band and time interval. To address this issue, sub-band CSP (SBCSP) was proposed by splitting raw EEG data into multi-level sub-bands in frequency domain (Novi et al., 2007). However, SBCSP ignores the relevance of different sub-bands, leading to the development of filter bank CSP (FBCSP) that incorporates a feature selection process based on mutual information

* Corresponding author.
E-mail address: wdzhou@sdu.edu.cn (W. Zhou).

https://doi.org/10.1016/j.jneumeth.2023.109953
Received 8 April 2023; Received in revised form 20 July 2023; Accepted 19 August 2023
Available online 21 August 2023

criteria (Ang et al., 2008). After acquiring manual features, classifiers such as linear discriminant analysis (LDA) (Gaur et al., 2021; Liu et al., 2022a), support vector machines (SVM) (Jin et al., 2019; Liu et al., 2022b), logistic regression (Ilyas et al., 2017), decision tree (DT) (Nisar et al., 2022), and K-nearest neighbors (KNN) (Baig et al., 2017) are typically employed to identify the type of MI tasks based on extracted features.

The performance of traditional methods relies highly on the quality of hand-crafted features. As a result, deep learning (DL) approaches (Al-Saegh et al., 2021; Craik et al., 2019) have become mainstream in MI tasks due to their advantage in end-to-end classification without manual intervention. Schirrmeister et al (Schirrmeister et al., 2017). proposed the Deep ConvNet and Shallow ConvNet models specifically designed for decoding EEG data, achieving performance on par with FBCSP by leveraging EEG temporal and spatial information. EEG-Net (Lawhern et al., 2018) was developed from the above ConvNet models and became the most commonly used DL approach for EEG-based BCIs. The depthwise and separable convolutions of EEG-Net were applied to extract temporal and spatial features from EEG signals and exhibited excellent performance on multiple BCI tasks such as MI and P300 paradigms. Yang et al (Yang et al., 2022). utilized a two-branch convolutional neural network employing continuous wavelet transform to learn the frequency features in one branch and the CNN module to learn the temporal features in another. This framework effectively captures temporal and frequency features, and improves the accuracy of MI-EEG decoding. Liu et al (Liu et al., 2022a). proposed SincNet-based hybrid neural network which uses SincNet to automatically identify the optimal cut-off frequency band for band-pass filters with EEG CSP features inputted into CNN and gated recurrent unit for classification. Roy (Roy, 2022) presented a multiscale CNN model which implemented multiscale convolution on each frequency band to extract semantic features from different frequency bands, thus achieving competitive performance in MI classification tasks.

While deep learning (DL) techniques to some extent alleviate some limitations of traditional methods associated with hand-crafted features, and enhance performances through automation, they suffer from its shortcomings in model interpretability and robustness (Liu et al., 2021, 2023b). The attention module was proposed to combine with a neural network to capture information from specific regions of images adaptively (Mnih et al., 2014; Bhattacharya et al., 2022). Recently, it started to be utilized in the field of MI EEG decoding. Liu et al (Liu et al., 2020). presented the spatial-temporal self-attention CNN-based architecture, where self-attention was utilized to capture the distinguishable spatial-temporal features and subsequently improve the accuracy of classification. Ma et al. (2022) proposed a long short-term memory (LSTM) approach combined with time-distributed attention, incorporating class attention mechanism and frequency band attention mechanism. This approach could adaptively assign different weights to various classes and frequency bands to achieve better performance in the 5-class MI scenario. Furthermore, a multi-branch EEG-Net with squeeze-and-excitation (SE) attention was applied to reduce the number of hyperparameters while increasing EEG MI classification accuracy (Altuwaijri, 15 et al., 2022). The promising results of attention-based approaches demonstrate the great potential of attention mechanisms in MI-EEG decoding.

Inspired by the attention model, self-attention-based CNN is proposed in this study to improve the ability of spatial features representation. Besides, time-frequency common spatial pattern (TFCSP) is presented to further extract the time-frequency-space information from EEG data, enhancing the MI-EEG classification performance. TFCSP is developed from traditional CSP and aims to address the shortcomings of traditional CSP that disregards time and frequency information. Features obtained by TFCSP and self-attention-based CNN are concatenated for classification using a fully connected layer. The combination of self-attention-based CNN and TFCSP makes our model extract features in time, frequency and spatial domains from EEG signals, thus enhancing the MI EEG classification performance. The proposed approach is evaluated using two MI datasets, with ablation experiments and parameter visualization demonstrating the effectiveness of the attention module and TFCSP algorithm. This work expands upon a conference paper presented at the 22nd International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS Conference 2022) (Zhang et al., 2022).

The remainder of this paper is organized as follows: Section II outlines the datasets utilized in this work, while Section III illustrates the details of the proposed methods. Section IV presents the experimental results, Section V discusses the findings, and Section VI draws conclusions.

2. EEG database

Two publicly available datasets, BCI Competition IV dataset IIa (BCI IV IIa) (Blanchard and Blankertz, 2004) and BCI Competition III dataset IIIa (BCI III IIIa) (Blankertz et al., 2006), are adopted to evaluate the proposed method. Both datasets have a sampled rate of 250 Hz, with differences in time scheme, number of channels, subjects, and trials. The time schemes for both datasets are provided in Fig. 1, with detailed information presented below.

The BCI IV IIa dataset was recorded from nine subjects (A01-A09) using 25 channels. The proposed method considers only the first 22 EEG channels, as the remaining three channels are monopolar electrooculogram channels. For each subject, two sessions of four MI tasks (left hand, right hand, both feet, and tongue) were conducted on two different days, with each session containing 288 trials (72 trials for each class). In this paper, the first session is used as the training set, while the other session is employed as the test set.

The BCI III IIIa dataset was recorded from three subjects (k3b, k6b, l1b) using 60 channels. Each subject completed four types of MI tasks (left hand, right hand, both feet, and tongue), with the subject containing 360 trials, 240 trials, and 240 trials, respectively. In this work, the total trials are evenly divided into a training set and a test set for each subject.

In our experiment, the EEG data of BCI IV IIa dataset from t = 2 s to t = 6 s and the EEG data of BCI III IIIa dataset from t = 3–7 s have been adopted for analysis.

3. Methods

The block diagram of the proposed self-attention-based CNN combined with TFCSP is shown in Fig. 2. The algorithm can be roughly divided into four steps: data augmentation (DA), TFCSP features

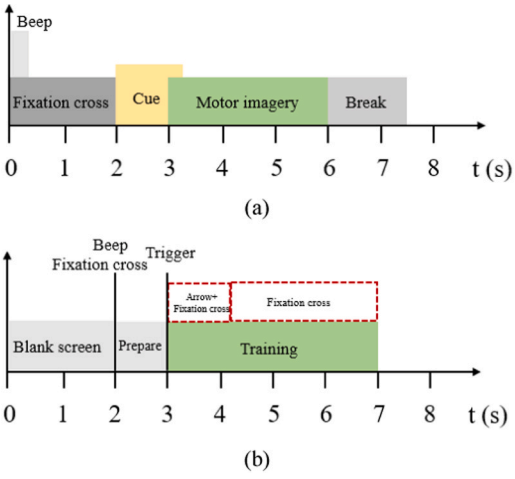

Fig. 1. The timing scheme of two MI datasets. (a) The paradigm of BCI IV IIa dataset. (b) The paradigm of BCI III IIIa dataset.

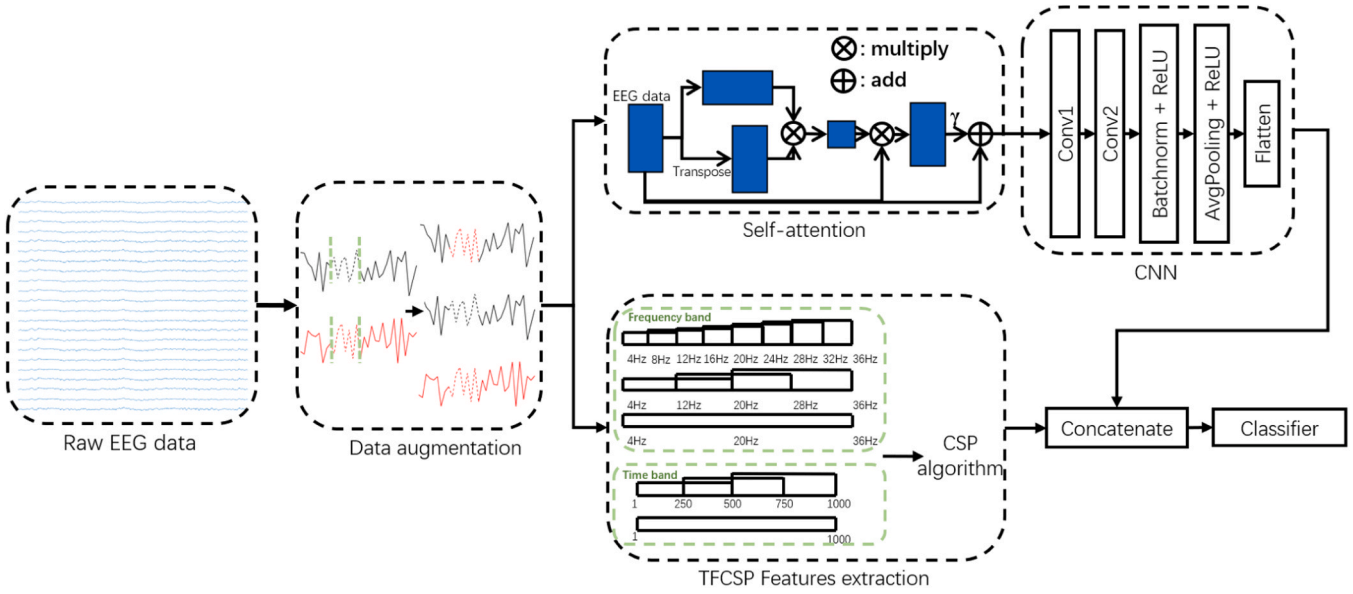

Fig. 2. Schematic illustration of the proposed self-attention-based CNN combining with TFCSP.

extraction, self-attention-based CNN, and classifier. Specifically, DA strategy is employed to expand the EEG dataset before processing the raw EEG data. Subsequently, TFCSP features are extracted from multiple time and frequency EEG bands. In the step of self-attention-based CNN, the self-attention module is utilized to yield channel weights which can strengthen the spatial information of EEG. Next, the CNN module is combined with self-attention to extract the temporal-spatial features. Finally, features acquired from the self-attention-based CNN module are concatenated with the TFCSP features for classification using a classifier composed of a softmax layer.

3.1. Data augmentation

The performance of CNN model is closely related to the amount of training dataset, and the scale of MI-based EEG database is typically small. Thus, a data augmentation method is proposed to address data insufficiency and improve classification accuracy. The DA procedure is shown in Fig. 3. The raw EEG signal (1000 time points for 4 s data) is divided into four consecutive and non-overlapping segments with 250 time points per segment. Different segments of EEG trials from the same class are regrouped randomly to yield new EEG epochs. The time order of the segments in one regrouped epoch remains consistent. Of note, exchanging the second segment yields the best performance compared to exchanging the third segment or exchanging the second and third segments simultaneously. The reason for the above phenomena may be that the second EEG segment, which is located in the middle period of motor imagery, is relatively stable and contains more information.

3.2. Time-frequency common spatial pattern (TFCSP)

Although the CSP algorithm is commonly used to extract spatial features from EEG, traditional CSP primarily focuses on EEG spatial features and disregards EEG time and frequency information. Therefore, time-frequency common spatial pattern (TFCSP) method is designed in this work. Raw EEG data is divided into time and frequency bands before extracting CSP features. Details of the dividing strategy in the frequency domain and the time domain are illustrated as Fig. 4(a) and (b), respectively. The frequency band between 4 and 36 Hz is partitioned into 11 subbands by the Butterworth filters with five orders. Under the bandwidth scales of 8 Hz, 16 Hz, 32 Hz, seven frequency bands, three subbands, and one band are acquired with 50 % frequency overlapping, respectively.

The division scheme in the time domain is similar to that in the frequency domain. Three time segments with a length of 500 time points, and one time segment with a length of 1000 time points are acquired with 50 % time overlapping. As a result, forty-four subbands

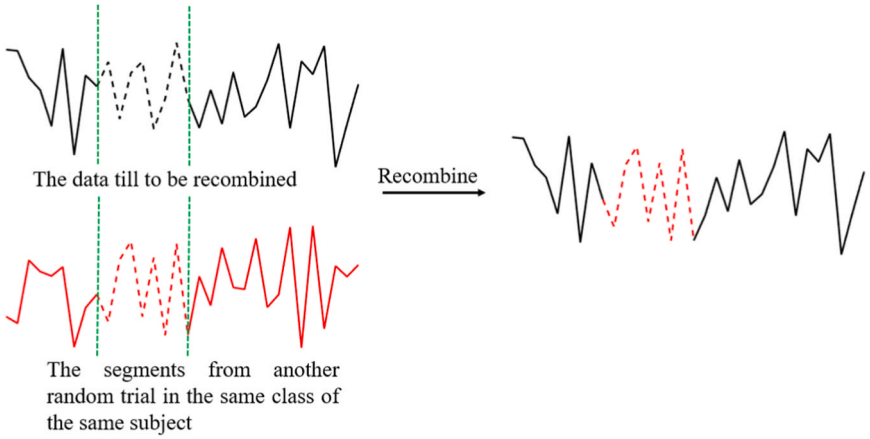

Fig. 3. Procedure of data reconstruction in time domain to expand data.

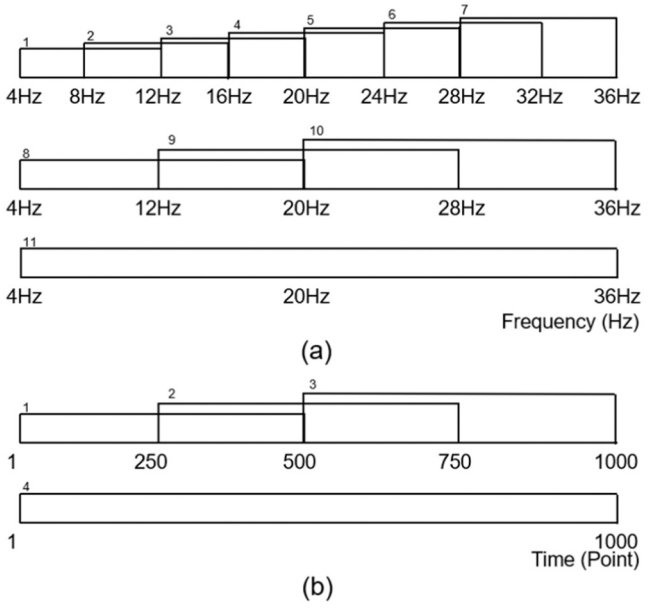

Fig. 4. Schematic representation of data division into time-frequency bands. The division strategy for the frequency domain is shown in (a), while the strategy for the time domain is illustrated in (b).

(11 frequency bands × 4 time bands) can be totally obtained in time and frequency domain.

After division, the CSP algorithm using One-Versus-One (OVO) strategy is applied to extract spatial features for each subband. In the context of multi-class MI tasks, OVO CSP obtains one CSP projection matrix by utilizing EEG signals for two different classes. This approach yields six spatial filters for four-class MI tasks. The steps of acquiring OVO CSP features are as follows:

(1) Calculating the normalized covariance matrix $R_{c,i}$ of the i-th trial belonging to the c-th class:

$$R_{c,i} = \frac{X_{c,i}X_{c,i}^{T}}{trace\left(X_{c,i}X_{c,i}^{T}\right)} \tag{1}$$

where $X_{c,i}$ denotes the EEG segments of the i-th trial belonging to the c-th class and trace(X) is the sum of the elements on the diagonal of X.

(2) Calculating the mixed space covariance matrix R:

$$\overline{R_c} = \frac{1}{k}\sum_{i=1}^{k} R_{c,i} \tag{2}$$

$$\overline{R_c} = \frac{1}{k}\sum_{i=1}^{k} R_{c,i} \tag{3}$$

$$R = \frac{\overline{R_1} + \overline{R_2}}{2} \tag{4}$$

where $\overline{R_c}$ denotes the mean covariance matrixes of k trials from the c-th class.

(3) Utilizing the whitening matrix P and Eq. (5) to acquire the spatial filter $W \epsilon R^{ch\times ch}$:

$$P = \frac{U_c^{T}}{\sqrt{\lambda_c}} \tag{5}$$

$$S_c = PR_cP^{T} = B_c\lambda_cB_c^{T} \tag{6}$$

where $B_C = U_c$.

$$W = B^{T}P \tag{7}$$

where the U_c and λ_C denote the eigenmatrix and the eigenvector of mixed space covariance matrix R, respectively. c represents the class of tasks.

(4) Obtaining the CSP features from EEG data by the projection of spatial filter $\omega \in R^{ch\times 2m}$, where ω is constructed by the first m rows and the last m rows of spatial filter W.

$$Z_{c,i} = \omega \times X_{c,i} \tag{8}$$

$$f_{c,i}^{p} = \log\left(\frac{\mathrm{var}\left(Z_{c,i}^{p}\right)}{\sum_{p=1}^{2m}\mathrm{var}\left(Z_{c,i}^{p}\right)}\right) \tag{9}$$

where $f_{c,i}^{p}$ is the p-th feature of the i-th trial belonging to the c-th class.

The value of m is set as 2, and the number of features in each subband is 24 (4 features × six spatial filters). The size of the feature input is 44 × 24 (the number of subbands × the number of CSP features for each band). The CSP features are flattened and concatenated with the features from self-attention-based CNN for further classifying.

3.3. Self-attention-based convolutional neural network (CNN)

The brain regions activated during the same MI task can differ among individuals, indicating that the motor-dependent channels are subject-specific. Traditional CNN lacks the ability to select EEG channels, resulting in suboptimal classification accuracy. Self-attention module is utilized to automatically identify the most useful channels, strengthen the weights of those channels in EEG, and improve the interpretability of the deep neural network model. The schematic diagram of self-attention module is illustrated in Fig. 5, where ch represents the number of channels (22 in the BCI IV IIa dataset and 60 in the BCI III IIIa dataset) and t (1000 in both datasets) refers to the time points of 4 s data. Raw EEG data of size $ch \times t$ is utilized to compute the query vector (Q), key vector (K), and value vector (V), with ch denoting the number of channels and t representing the time points. The vectors of Q, K^T, and V are equal to the raw EEG matrix. $Q = \{q_1, q_2, \ldots, q_{ch}\}$ and $K = \{k_1, k_2, \ldots, k_{ch}\}$ are utilized to achieve the attention weights matrix $P \in R^{ch\times ch}$. The detailed calculating steps are as follows:

$$S = Q \times K^{T} = \{q_1, q_2, \ldots, q_{ch}\} \times \left\{k_1^{T}, k_2^{T}, \ldots, k_{ch}^{T}\right\} \tag{10}$$

$$P = \mathrm{Softmax}\left(\frac{S}{\mathrm{std}(S)}\right) \tag{11}$$

where $S \in R^{ch\times ch}$ denotes the attention weights matrix before being scaled by the softmax function, and std(X) is the function of calculating standard deviation.

After acquiring the attention weights matrix, the output $O \in R^{ch\times t}$ is calculated by multiplying P and $V \in R^{ch\times t}$:

$$\begin{aligned} O = P \times V &= \begin{pmatrix} p_{1,1} & \cdots & p_{1,ch} \\ \vdots & \ddots & \vdots \\ p_{ch,1} & \cdots & p_{ch,ch} \end{pmatrix} \times \begin{pmatrix} v_1 \\ \vdots \\ v_{ch} \end{pmatrix} \\ &= \begin{pmatrix} p_{1,1}v_1 + \cdots + p_{1,ch}v_{ch} \\ \vdots \\ p_{ch,1}v_1 + \cdots + p_{ch,ch}v_{ch} \end{pmatrix} \end{aligned} \tag{12}$$

The output matrix encapsulates the mutual information between different channels. As depicted in Fig. 2, the self-attention module computes the output matrix, multiplies it by the learnable parameter $\gamma \in R^{ch\times t}$, and adds it to the raw EEG matrix. This process enables γ to automatically select the active channels and strengthen their weights.

The self-attention-based CNN comprises a self-attention layer, two

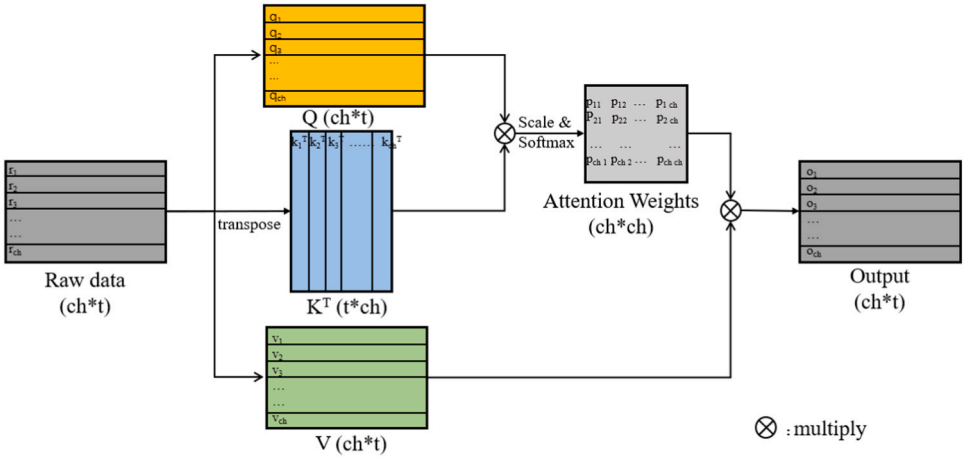

Fig. 5. The schematic diagram of the self-attention module.

convolutional layers (Conv1 and Conv2), a batch normalization layer (BatchNorm), a non-linear activation layer (ReLU), and an average pooling layer (AvgPooling). The detailed architecture of self-attention-based CNN is illustrated in Table 1. A self-attention layer is designed to strengthen channel weights of raw EEG data. The first convolutional layer (Conv1) with 40 filters of size 1×25 is employed to extract the temporal information. The second convolutional layer (Conv2) with 40 filters of size $ch \times 1$ is utilized to acquire spatial information. The shape of the output is transformed from (*ch*, 1000, 1) to (1, 976, 40). Batch normalization is used to accelerate the training process and avoids the gradient vanishing issue. The ReLU layer serves as a non-linear activation function. The average pooling layer, with a kernel size of 1×75 and a stride of 1×15, downsamples and reduces data redundancy. Features derived from self-attention-based CNN are flattened and concatenated with TFCSP features for final classification using a softmax layer consisting of four units.

4. Results

All experiments are implemented and tested using MATLAB R2022a on an Intel i9–9820X CPU, 64 GB RAM, and Nvidia RTX 2080 Super platform. Two datasets, BCI III IIIa and BCI IV IIa are used to validate the performance of the proposed approach, and classification accuracy serves as the evaluation metric and is calculated as follows:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \times 100\% \tag{13}$$

where *TP* is the number of true positives, *TN* is the number of true negatives, *FP* is the number of false positives, and *FN* is the number of false negatives.

The results of traditional DL methods are displayed in the first four rows of Table II and Table III, while the outcomes of machine learning approaches are listed in the fifth and sixth rows. As shown in Table II, our approach outperforms EEGNet (Lawhern et al., 2018), ConvNet (Schirrmeister et al., 2017) and ST-Attention CNN (Liu et al., 2020), achieving average classification accuracy improvements of 26.11 %, 19.54 %, 49.35 %, 27.31 % on the BCI III IIIa dataset. In comparison with traditional machine learning methods, such as FBCSP (Ang et al., 2008) and the multiscale time-frequency method (Liu et al., 2022a), our approach achieves the highest classification accuracy for the majority of subjects in the BCI III IIIa dataset. For subjects 'k6b' and 'l1b' in BCI III IIIa dataset, our method attains the optimal classification accuracies, whereas for the subject 'k3b', our approach's classification accuracy of is 1.67 % lower than the best-performing approach, the multiscale time-frequency method (Liu et al., 2022a). Table III reveals that the proposed method outperforms other approaches on the BCI IV IIa dataset with an average accuracy of 79.28 %, and achieves the highest classification accuracy for each subject in the dataset.

Although DL approaches can extract spatial-temporal features through time convolution and spatial filtering, our method achieves better classification accuracy, this may be attributed to our method's superiority in extracting spatial information compared to traditional CNN models. Furthermore, Deep ConvNet (Schirrmeister et al., 2017), which comprises four convolution-pooling block modules, yields the lowest accuracy, possibly due to overfitting in small datasets. For the machine learning methods, the average accuracy of the proposed method is the highest, which may be because CSP-based machine learning methods ignore temporal information of EEG.

In addition, the results of Table 2 and Table 3 revealed significant variability in the accuracy across subjects. High accuracy in MI-EEG decoding was exhibited in some individuals, whereas a few cases of lower accuracy results were also observed. Such individual differences may mainly be caused by the non-stationary nature of EEG signals and the existent of artifacts.

The total average inference time of our model for each trial in BCI III IIIa is around 429.54 ms, where TFCSP feature extraction time is 426.40 ms and self-attention-based CNN time is 3.14 ms. In addition, the average inference time for each trial in BCI IV IIa is about 110.53 ms, where TFCSP feature extraction and self-attention-based CNN take around 108.74 ms and 1.79 ms, respectively.

In order to further verify the validity of each module in this paper, ablation experiments are conducted. As shown in Fig. 6(a), TFCSP outperforms both the self-attention-based CNN and the combination of CSP and self-attention-based CNN on the BCI IV IIa dataset, achieving mean

Table 1
Detailed Architecture of Self-attention-based CNN.

Layer	Self-attention	Conv1	Conv2	BatchNorm	ReLU	AvgPooling	ReLU	Flatten
Input	(*ch*, 1000, 1)	(*ch*, 1000, 1)	(*ch*, 976, 40)	(1, 976, 40)	(1, 976, 40)	(1, 976, 40)	(1, 61, 40)	(1, 61, 40)
Output	(*ch*, 1000, 1)	(*ch*, 976, 40)	(1, 976, 40)	(1, 976, 40)	(1, 976, 40)	(1, 61, 40)	(1, 61, 40)	(1, 1, 2440)
Feature maps	1	40	40	40	40	40	40	2440
Kernel	–	(1, 25)	(*ch*, 1)	–	–	(1, 75)	–	–
Stride	–	(1, 1)	(1, 1)	–	–	(1, 15)	–	–

Table 2
The Accuracy (%) Comparison on BCI IV IIa Dataset.

Method	Subjects									Mean
	A01	A02	A03	A04	A05	A06	A07	A08	A09	
EEGNet (Schirrmeister et al., 2017)	78.82	53.82	80.90	61.11	68.75	58.68	73.61	75.35	67.71	68.75
ST-Attention CNN (Liu et al., 2023b)	72.57	54.86	78.47	55.90	69.10	50.00	69.79	67.71	68.06	65.16
Deep ConvNet (Craik et al., 2019)	48.96	44.79	52.78	43.40	46.53	35.07	54.86	44.79	42.36	45.95
Shallow ConvNet (Craik et al., 2019)	71.53	52.08	82.64	63.54	73.61	53.47	72.57	73.61	72.22	68.36
FBCSP (Novi et al., 2007)	78.13	49.65	76.74	60.42	57.29	45.14	81.60	76.74	65.28	65.66
Multiscale time-frequency method (Gaur et al., 2021)	84.03	61.81	82.99	63.89	70.14	52.08	92.01	81.25	80.21	74.27
Proposed Method	**85.76**	**62.50**	**87.15**	**76.04**	**78.82**	**59.72**	**92.36**	**86.46**	**84.72**	**79.28**

Table 3
The Accuracy (%) Comparison on BCI III IIIa Dataset.

Method	Subjects			Mean
	K3b	K6b	L1b	
EEGNet (Schirrmeister et al., 2017)	70.00	59.17	51.67	60.28
ST-Attention CNN (Liu et al., 2023b)	80.56	64.67	58.33	66.85
Deep ConvNet (Craik et al., 2019)	44.44	35.83	30.83	37.04
Shallow ConvNet (Craik et al., 2019)	63.89	64.17	49.17	59.08
FBCSP (Novi et al., 2007)	95.00	72.50	65.00	77.50
Multiscale time-frequency method (Gaur et al., 2021)	**96.67**	71.67	80.00	82.78
Proposed Method	95.00	**80.83**	**83.33**	**86.39**

accuracy improvements of 1.78 % and 0.89 %, respectively. Fig. 6(b) shows that, for the BCI III IIIa dataset, TFCSP's mean accuracy is 16.39 % higher than that of the self-attention-based CNN and 17.41 % higher than that of the combination of CSP and self-attention-based CNN. Moreover, as shown in Fig. 6(a) and (b), the self-attention-based CNN combined with TFCSP also enhances classification accuracies for both BCI IV IIa and BCI III IIIa datasets. Compared to TFCSP alone, the combination of self-attention-based CNN and TFCSP further improves the classification performance, achieving mean classification accuracy of 78.13 % on BCI IV IIa and 85.46 % on BCI III IIIa. Fig. 6 reveals that DA module contributes to the increased mean accuracy (0.53 % improvement on BCI III IIIa and 1.15 % improvement on BCI IV IIa) due to its capacity to generate additional EEG data and expand datasets scales.

Moreover, the topographical maps of the learnable parameter γ in the self-attention module are depicted in Fig. 7 to further demonstrate the effectiveness of the self-attention module, highlighting specific activated brain regions during MI and effectively displaying the differences in active brain regions and motor-dependent EEG channels for various subjects.

5. Discussions

Traditional CSP algorithm focused on EEG spatial features, ignoring EEG time and frequency information, and their performance depended on the selected frequency band. SBCSP and FBCSP were developed from CSP algorithm by splitting raw EEG data into multi-level sub-bands in frequency domain, but neglecting the information in time domain and the correlation between information in frequency domain. The proposed TFCSP segments raw EEG data into multiple time-frequency bands before projecting it into CSP space, addressing the issue of CSP neglecting EEG time and frequency information. TFCSP can take full advantage of EEG time-frequency information by employing the multi-scale division strategy in time and frequency domains. The comparative results in Fig. 6 highlight the efficacy of TFCSP features.

The results in Fig. 6 demonstrate that the self-attention module improves its ability in extracting spatial information. In contrast to traditional CNN, self-attention module captures potential spatial links

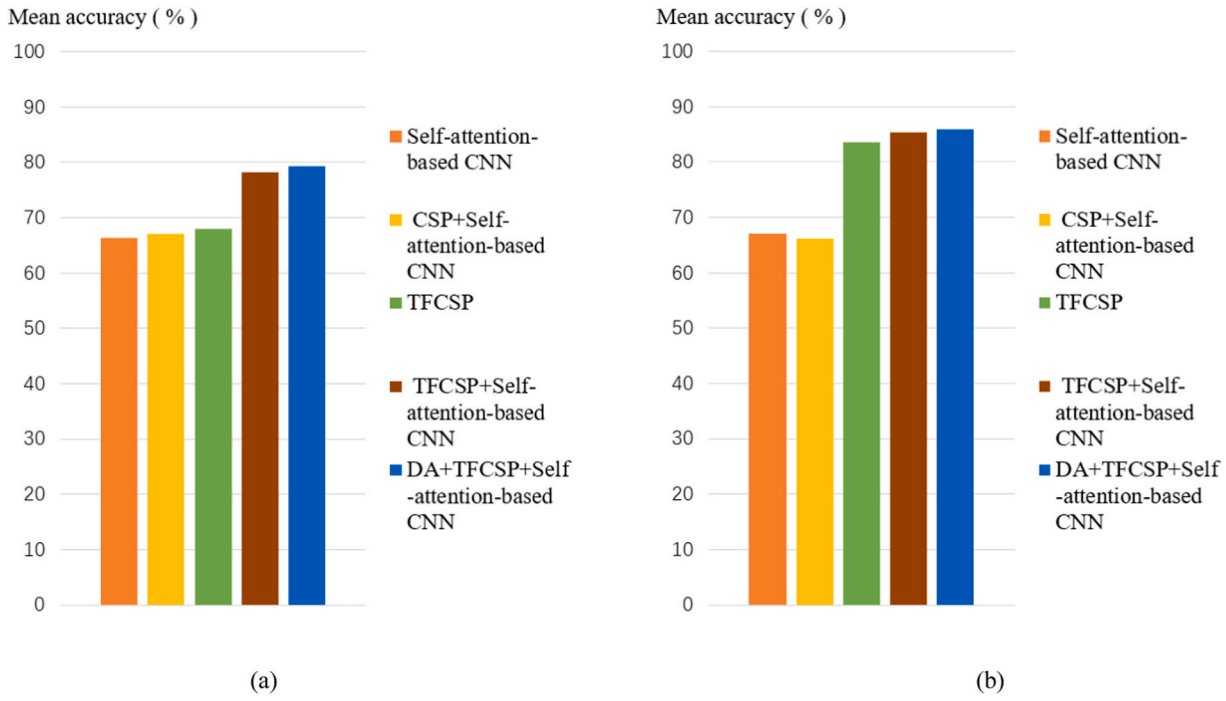

Fig. 6. The comparison accuracy results of ablation experiments on dataset BCI IV IIa and dataset BCI III IIIa are shown in (a) and (b), respectively.

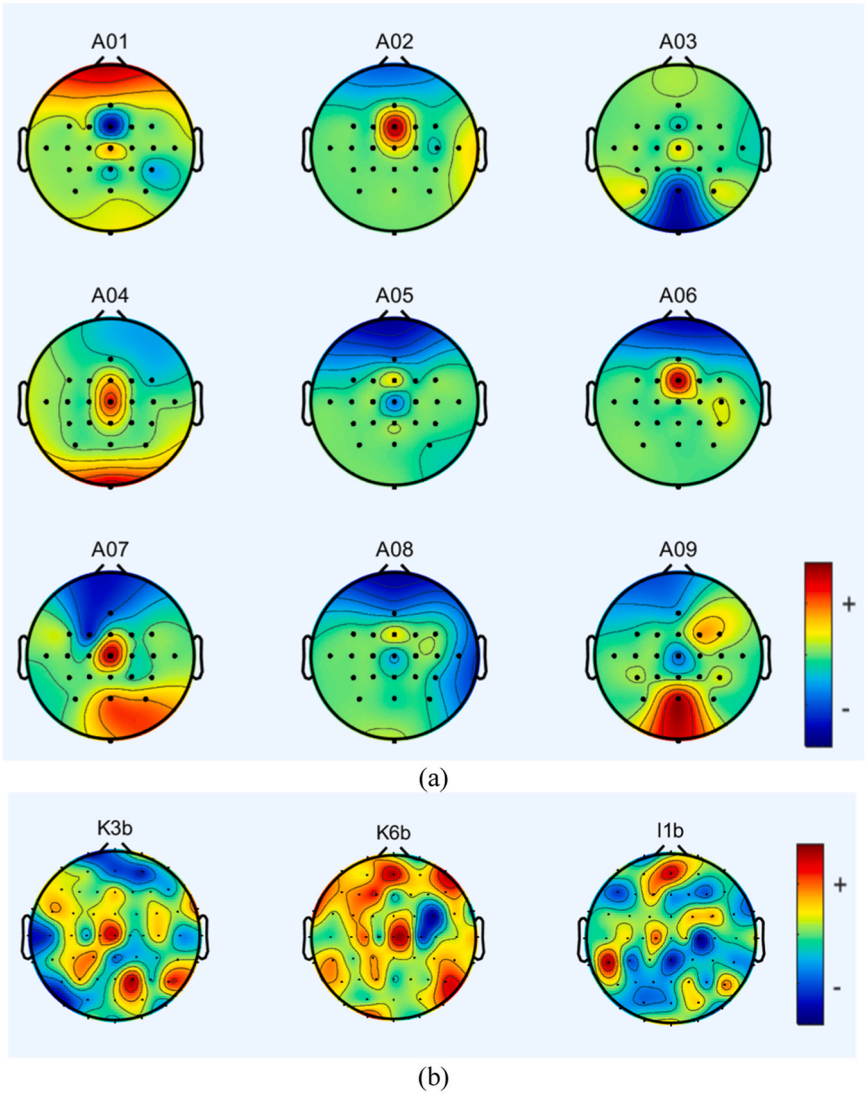

Fig. 7. The topographical maps of the learnable parameter in the self-attention module. (a) The topographical maps for nine subjects from BCI IV IIa dataset. (b) The topographical maps for three subjects from BCI III IIIa dataset.

between channels and enhances the interpretability of model. As shown in Fig. 7, the learnable parameter γ in the proposed self-attention module, which represents the EEG channel weights of different subjects, can automatically select active channels by assigning higher values to motor-related channels. The differences in channel weights demonstrate that active EEG channels during motor imagery are different between diverse subjects. Therefore, the MI classification accuracies could be enhanced by increasing the weights of motor-dependent channels for different individuals.

Though the proposed approach achieves promising performance, there are still some limitations in our method. First, compared with CSP method, TFCSP increases the computational complexity. Second, the experiment is conducted on subject-dependent evaluation. In addition, our method is only trained and tested on offline BCI datasets. In our future work, we will further optimize the proposed TFCSP to reduce its computational cost and enhance its adaptability for online BCI applications. Besides, subject-independent experiment will be conducted to substantiate the generality and robustness of our model across diverse subjects.

6. Conclusions

EEG-based MI BCI is a promising technology due to its broad applicability in both medical and non-medical fields. However, decoding MI EEG signals remains challenging due to individual variations in active brain regions during MI tasks, leading to suboptimal performance at odds with practical application. In this work, a self-attention-based CNN combined with TFCSP is proposed to classify EEG signals in four-class MI tasks. The self-attention-based CNN is trained to automatically obtain the temporal and spatial information from EEG signals, while the self-attention module characterizes the active channels by quantitating EEG channel weights. Moreover, TFCSP is introduced to further extract the time-frequency-space information of EEG in a multi-scale approach. The integration of TFCSP and the self-attention-based CNN leverages the time-frequency-space information of EEG, and thus enhances the classification performance. Our method is evaluated on two datasets, BCI IV IIa and BCI III IIIa, achieving outstanding accuracies of 79.28 % and 86.39 %, respectively, outperforming other state-of-the-art MI classification methods. Furthermore, the ablation experiments also substantiate the effectiveness of both TFCSP and the self-attention-based CNN.

CRediT authorship contribution statement

Rui Zhang: Conceptualization, Data curation, Formal analysis, Methodology, Software, Validation, Writing - original draft. **Guoyang Liu:** Conceptualization, Methodology, Writing - review & editing. **Yiming Wen:** Formal analysis, Writing - review & editing. **Weidong Zhou:** Conceptualization, Funding acquisition, Supervision, Writing - review & editing.

Declaration of Competing Interest

The authors declared no potential conflicts of interest with respect to the research, author- ship, and/or publication of this article.

Data availability

The data that has been used is confidential.

Acknowledgements

The support of National Natural Science Foundation of China (No. 62271291), the Key Program of the Natural Science Foundation of Shandong Province under Grant ZR2020LZH009, and the Research Funds of Science and Technology Innovation Committee of Shenzhen Municipality under Grant JCYJ20180305164357463 is gratefully acknowledged.

References

Abiri, R., Borhani, S., Sellers, E.W., Jiang, Y., Zhao, X., 2019. A comprehensive review of EEG-based brain-computer interface paradigms. J. Neural Eng. 16 (1), 011001.
Al-Quraishi, M.S., Elamvazuthi, I., Daud, S.A., Parasuraman, S., Borboni, A., 2018. EEG-based control for upper and lower limb exoskeletons and prostheses: a systematic review, (in eng). Sens. (Basel, Switz.) 18 (10).
Al-Saegh, A., Dawwd, S.A., Abdul-Jabbar, J.M., 2021. Deep learning for motor imagery EEG-based classification: a review. Biomed. Signal Process. Control 63.
Altuwaijri, G.A., Muhammad, G., Altaheri, H., Alsulaiman, M., 2022. A multi-branch convolutional neural network with squeeze-and-excitation attention blocks for EEG-based motor imagery signals classification. Diagn. (Basel) 12 (4).
Ang, K.K., Chin, Z.Y., Zhang, H., Guan, C., 2008. Filter bank common spatial pattern (FBCSP) in brain-computer interface. Proc. Int. Jt. Conf. Neural Netw. 2390–2397.
Baig, M.Z., Aslam, N., Shum, H.P.H., Zhang, L., 2017. Differential evolution algorithm as a tool for optimal feature subset selection in motor imagery EEG, 2017/12/30/ Expert Syst. Appl. 90, 184–195, 2017/12/30/.
Bhattacharya, A., Baweja, T., Karri, S.P.K., 2022. Epileptic seizure prediction using deep transformer model. Int J. Neural Syst. 32 (2), 2150058.
Blanchard, G., Blankertz, B., 2004. BCI competition 2003-data set IIa: spatial patterns of self-controlled brain rhythm modulations. IEEE Trans. Biomed. Eng. 51 (6), 1062–1066.
Blankertz, B., et al., 2006. The BCI competition III: validating alternative approaches to actual BCI problems. IEEE Trans. Neural Syst. Rehabil. Eng. 14 (2), 153–159.
Craik, A., He, Y., Contreras-Vidal, J.L., 2019. Deep learning for electroencephalogram (EEG) classification tasks: a review. J. Neural Eng. 16 (3), 031001.
Elstob, D., Lindo Secco, E., 2016. A low cost eeg based Bci prosthetic using motor imagery. Int. J. Inf. Technol. Converg. Serv. 6 (1), 23–36.
Gaur, P., Gupta, H., Chowdhury, A., McCreadie, K., Pachori, R.B., Wang, H., 2021. A sliding window common spatial pattern for enhancing motor imagery classification in EEG-BCI. IEEE Trans. Instrum. Meas. 70, 1–9.
Graimann, B., Allison, B., Pfurtscheller, G., 2010. Brain–computer interfaces: a gentle introduction. In: Graimann, B., Pfurtscheller, G., Allison, B. (Eds.), Brain-Computer Interfaces: Revolutionizing Human-Computer Interaction. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 1–27.
Ieracitano, C., Morabito, F.C., Hussain, A., Mammone, N., 2021. A hybrid-domain deep learning-based BCI for discriminating hand motion planning from EEG sources (Sep). Int J. Neural Syst. 31 (9), 2150038.
Ilyas, M.Z., Saad, P., Ahmad, M.I., Ghani, A., 2017. Classification of EEG signals for brain-computer interface applications: performance comparison. 2016 Int. Conf. Robot. Autom. Sci. (ICORAS).
Jin, J., Miao, Y., Daly, I., Zuo, C., Hu, D., Cichocki, A., 2019. Correlation-based channel selection and regularized feature optimization for MI-based BCI. Neural Netw. 118, 262–270, 2019/10/01/.
Khan, M.A., Das, R., Iversen, H.K., Puthusserypady, S., 2020. Review on motor imagery based BCI systems for upper limb post-stroke neurorehabilitation: from designing to application. Comput. Biol. Med. 123, 103843.
Koles, Z.J., Lazar, M.S., Zhou, S.Z., 1990. Spatial patterns underlying population differences in the background EEG, (in eng). Brain Topogr. 2 (4), 275–284.
LaFleur, K., Cassady, K., Doud, A., Shades, K., Rogin, E., He, B., 2013. Quadcopter control in three-dimensional space using a noninvasive motor imagery-based brain-computer interface. J. Neural Eng. 10 (4), 046003.
Lawhern, V.J., Solon, A.J., Waytowich, N.R., Gordon, S.M., Hung, C.P., Lance, B.J., 2018. EEGNet: a compact convolutional neural network for EEG-based brain-computer interfaces. J. Neural Eng. 15 (5), 056013.
Liu, C., et al., 2022a. SincNet-based hybrid neural network for motor imagery EEG decoding. IEEE Trans. Neural Syst. Rehabil. Eng. 30, 540–549.
Liu, G., Han, X., Tian, L., Zhou, W., Liu, H., 2021. ECG quality assessment based on hand-crafted statistics and deep-learned S-transform spectrogram features. Comput. Methods Prog. Biomed. 208, 106269.
Liu, G., Tian, L., Zhou, W., 2022a. Multiscale time-frequency method for multiclass motor imagery brain computer interface. Comput. Biol. Med. 143, 105299.
Liu, G., Hsiao, J., Zhou, W., Tian, L., 2023a. MartMi-BCI: a matlab-based real-time motor imagery brain-computer interface platform. SoftwareX 22, 101371.
Liu, G., Zhang, J., Chan, A.B., Hsiao, J., 2023b. Human attention-guided explainable AI for object detection. Proc. Annu. Meet. Cogn. Sci. Soc. 45 (45).
Liu, X., Shen, Y., Liu, J., Yang, J., Xiong, P., Lin, F., 2020. Parallel spatial-temporal self-attention CNN-based motor imagery classification for BCI, (in eng). Front Neurosci. 14, 587520.
Liu, Y., Wang, Z., Huang, S., Wang, W., Ming, D., 2022b. EEG characteristic investigation of the sixth-finger motor imagery and optimal channel selection for classification. J. Neural Eng. 1, 19.
Ma, X., Qiu, S., He, H., 2022. Time-distributed attention network for EEG-based motor imagery decoding from the same limb. IEEE Trans. Neural Syst. Rehabil. Eng. 30, 496–508.
V. Mnih, N. Heess, A. Graves, and K. Kavukcuoglu, Recurrent Models of Visual Attention, p. arXiv:1406.6247Accessed on: June 01, 2014[Online]. Available: ⟨https://ui.adsabs.harvard.edu/abs/2014arXiv1406.6247M⟩.
Nisar, H., Wee Boon, K., Kim Ho, Y., Shen Khang, T., 2022. Brain-computer interface: feature extraction and classification of motor imagery-based cognitive tasks, presented at the 2022. IEEE Int. Conf. Autom. Control Intell. Syst. (I2CACIS).
Novi, Q., Guan, C., Dat, T.H., Xue, P., 2007. Sub-band common spatial pattern (SBCSP) for brain-computer interface. 2007 3rd Int. IEEE/EMBS Conf. Neural Eng. 204–207.
Padfield, N., Zabalza, J., Zhao, H., Masero, V., Ren, J., 2019. EEG-based brain-computer interfaces using motor-imagery: techniques and challenges. Sens. (Basel, Switz.) 19 (6).
Pfurtscheller, G., Neuper, C., 2001. Motor imagery and direct brain-computer communication. Proc. IEEE 89 (7), 1123–1134.
Pfurtscheller, G., Neuper, C., Flotzinger, D., Pregenzer, M., 1997. EEG-based discrimination between imagination of right and left hand movement, (in eng). Electroencephalogr. Clin. Neurophysiol. 103 (6), 642–651.
Pfurtscheller, G., Brunner, C., Schlogl, A., Lopes da Silva, F.H., 2006. Mu rhythm (de) synchronization and EEG single-trial classification of different motor imagery tasks. Neuroimage 31 (1), 153–159.
Roy, A.M., 2022. An efficient multi-scale CNN model with intrinsic feature integration for motor imagery EEG subject classification in brain-machine interfaces. Biomed. Signal Process. Control 74.
Schirrmeister, R.T., et al., 2017. Deep learning with convolutional neural networks for EEG decoding and visualization. Hum. Brain Mapp. 38 (11), 5391–5420.
Xie, P., Hao, S., Zhao, J., Liang, Z., Li, X., 2022. A spatio-temporal method for extracting gamma-band features to enhance classification in a rapid serial visual presentation task. Int J. Neural Syst. 32 (3), 2250010.
Xu, F., Zhou, W., Zhen, Y., Yuan, Q., Wu, Q., 2016. Using fractal and local binary pattern features for classification of ECOG motor imagery tasks obtained from the right brain hemisphere. Int. J. Neural Syst. 26 (6), 1650022.
Yang, J., Gao, S., Shen, T., 2022. A two-branch CNN fusing temporal and frequency features for motor imagery EEG decoding. Mar 8 Entropy (Basel) vol. 24 (3). Mar 8.
Yu, Z., Chen, W., Zhang, T., 2022. Motor imagery EEG classification algorithm based on improved lightweight feature fusion network. Biomed. Signal Process. Control 75.
Zhang, R., et al., 2022. Motor imagery EEG classification with self-attention-based convolutional neural network. 2022 7th Int. Conf. Intell. Inform. Biomed. Sci. (ICIIBMS) 7, 195–199.
Zhang, X., Yao, L., Wang, X., Monaghan, J., McAlpine, D., Zhang, Y., 2021. A survey on deep learning-based non-invasive brain signals: recent advances and new frontiers. J. Neural Eng. 18 (3).